A Bayesian Network Schema for Lessening Database Inference

LiWu Chang and Ira S. Moskowitz
Center for High Assurance Computer Systems
Naval Research Laboratory
Washington, DC 20375
lchang@itd.nrl.navy.mil
in proceeding of CIMCA01

Abstract

"Database inference" occurs when unauthorized users infer sensitive information from publicly released data. To protect against such "inference attacks," information that is probabilistically related to sensitive information must be examined and perhaps modified. We introduce a formal schema for database inference analysis, based upon a Bayesian network structure, which identifies critical parameters involved in the inference problem and represents them in a coherent framework.

1 Introduction

"Database inference" occurs when unauthorized users infer sensitive information from publicly released data. To protect against such "inference attacks," information that is probabilistically related to sensitive information must be examined and perhaps modified. The typical analysis of the probabilistic dependency relationships is carried out using Bayesian network theory [1][2][3]. Pearl [1][4] has shown how Bayesian networks *model* inference. We use the same technique to *lessen* inference. Specifically, we introduce a formal schema for database inference analysis, based upon a Bayesian network structure, which identifies critical parameters involved in the inference problem and represents them in a coherent framework.

Although several researchers offer different approaches to mitigating the database inference problem, (e.g., [5][6][7][8][9][10][11]), we are the first to use a Bayesian network approach ([12][13]). (In this paper we do not discuss how to construct a Bayesian network B_n for a given database, see e.g., [3][14].) The most common technique for protecting sensitive information is that of downgrading the non-sensitive information in the database, also referred to as database *sanitization*. The result of downgrading is to mitigate, if not eradicate, the inference problem. We feel that it is important to describe the downgrading issue in terms of a Bayesian network because we see which (and how) attributes impact upon sensitive information. We describe our schema by a six tuple $<I, \tau, S, V, O, E>$, where

1. I (*input*): is a relational database;
2. τ (*tolerance*): is the measure of information loss that users are willing to tolerate in order to obtain to data protection;
3. S (*search strategy*): is the strategy for searching desired attribute values from the data set;

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2001		**00-00-2001 to 00-00-2001**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
A Bayesian Network Schema for Lessening Database Inference	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,Center for High Assurance Computer Systems,4555 Overlook Avenue, SW,Washington,DC,20375**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**7**	

Table 1: $D_H = I$ — *sample medical records*
(U: uid; H: hepatitis; D: depression; A: AIDS; T: low thyroid; F: transfusion)

U	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
H	n	y	y	y	n	n	y	y	y	n	n	y	n	y	n	y	n	y	n	y
D	n	y	y	y	y	n	n	n	y	n	y	n	n	y	y	n	y	y	y	n
A	n	n	y	y	n	n	n	y	y	n	y	n	n	y	n	n	n	y	n	n
T	n	y	n	y	n	n	n	y	n	n	y	n	n	y	y	n	y	n	y	n
F	n	n	y	n	n	n	n	y	n	n	y	n	n	n	n	n	n	y	n	y

4. V (*data selection criterion*): is the criterion for choosing attribute values to downgrade or modify;
5. O (*output*): is a set of selected attribute values; and
6. E (*post evaluation criterion*): is the criterion for measuring the effect of downgrading.

The terms in the six tuple are dependent upon the choice of Bayesian network B_n.

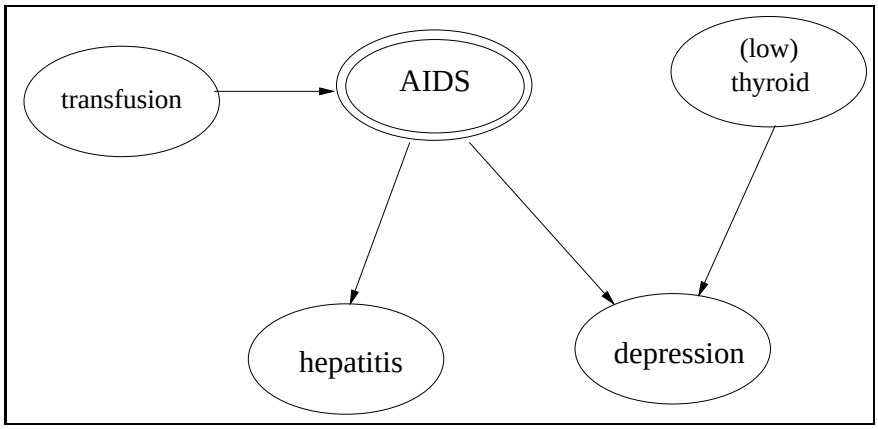

Figure 1: Architecture of a Bayesian Network (for High B_n^H). An attribute is denoted by a node. An arrow indicates the probabilistic dependency between two attributes. A double circle denotes that the attribute is sensitive.

We use the sample medical records shown in Table 1 as our example. We use High (H) (all the information) and Low (L) (the non-sensitive information) ([10]) to indicate, respectively, the portion of a database viewed by a database manager (the High user) and a generic (Low) user. Table 1 shows the High view (denoted here as D_H). D_H is I, the input database. A corresponding Bayesian network representation [3] is given in Figure 1, which shows that "AIDS" causes both "hepatitis" and "depression." Note that "depression" is also caused by (low) thyroid, and a cause of "AIDS" is a "transfusion." Here, the sensitive information is the diagnosis of "AIDS." Table 2 shows the database after being downgraded (denoted here as D_L). The dashes represent data that is considered sensitive and, thus, is not downgraded. A target node T is a node that has dashes in it (from Low's viewpoint). Thus, T represents sensitive information. We wish to lessen any inferences that a Low user may attempt to draw about the target node (sensitive information).

Table 2: D_L — *medical records of Low database*
(U: uid; H: hepatitis; D: depression; A: AIDS; T: low thyroid; F: transfusion)

U	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
H	n	y	y	y	n	n	y	y	y	n	n	y	n	y	n	y	n	y	n	y
D	n	y	y	y	y	n	n	n	y	n	y	n	n	y	y	n	y	y	y	n
A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T	n	y	n	y	n	n	n	y	n	n	y	n	n	y	y	n	y	n	y	n
F	n	n	y	n	n	n	n	y	n	n	y	n	n	n	n	n	n	y	n	y

Since data is not completely revealed, the corresponding Bayesian network structure for D_L differs from that of D_H and is shown in Figure 2. The challenge for a Low user who is attempting to discern sensitive information is to restore the downgraded information in Table 2. Note that Table 2 still contains the "AIDS" attribute, even though the values are all missing. This is because we take the paranoid view that Low knows what sensitive attribute High is concerned with, and because, in general, sensitive information may be distributed across many attributes and all the values may not be missing.

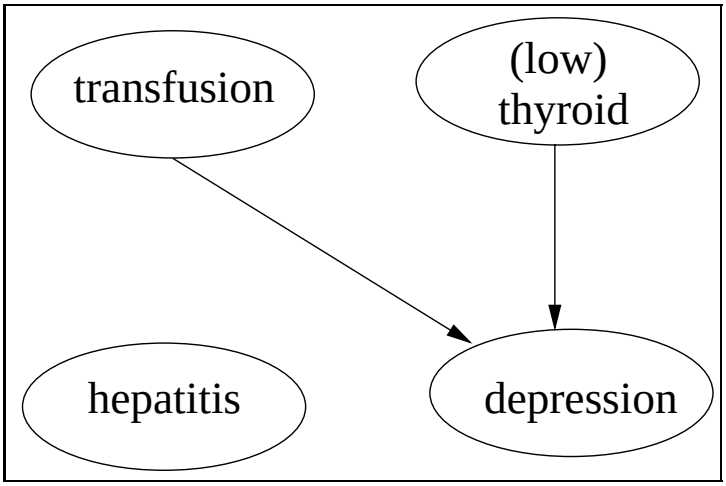

Figure 2: B_n^L The Bayesian Network from Low Database.

2 Information Reduction

Because certain non-sensitive information can lead to probabilistic inferences about the sensitive information, we approach the problem of lessening inference by not downgrading all of the non-sensitive information. Thus, with respect to a database protection strategy, the effective method of mitigating inference we propose is to modify non-sensitive data by "blocking," i.e., replacing an attribute value with a "?," indicating no knowledge about the attribute value. Given a database D, we let D^m denote D after at least one of its entries has been modified. We do not use imputation, which introduces erroneous data (i.e., replacing an attribute value with another different value), because of the negative performance side effects.

On the other hand, the less information blocked, the better, from the performance standpoint. Therefore, instead of sending D_L to Low, High instead blocks some of the non-sensitive information, and sends D_L^m to Low. We use the Bayesian network structure to intelligently perform modifications and block non-sensitive information. The Bayesian network contains the sensitive information in the target node (assume only one target node) and the graphical structure models causal inferences [1][4]. In the following sections, we discuss how to select desired attributes, and then how to select values for those attributes which we will modify.

3 τ - Tolerance

Our pragmatic policy of preventing/lessening inference states that modification of non-sensitive information should lessen inference of sensitive information, while at the same time minimizing loss of functionality. It is a challenge to respect these two competing goals! Since modification affects the functionality of the database, we use a metric τ to describe the change. The definition of τ uses the probabilistic term $Pr(D|B_n)$ [3], which describes how likely a database D is to have a given Bayesian network B_n.

$$\tau \stackrel{\text{def}}{=} \frac{|\log Pr(D_L|B_n^L) - \log Pr(D_L^m|B_n^{L*})|}{|\log Pr(D_L|B_n^L)|}$$

We compute the sample probability of the modified database, $Pr(D_L^m|B_n^{L*})$, as the average over all possible instantiations of those values, where B_n^{L*} indicates the different Bayesian networks that are induced by the different instantiations. The tolerance τ provides a margin within which the information protection strategies operate. Thus, we often associate an upper bound U to τ, so that $\tau \leq U$. (τ is different than the operation ratio given in [13], in that τ is a metric of the Low view.)

4 *S* - Search Strategy

The search — how does one decide what attribute values are to be modified to lessen inference? It is impractical to perform exhaustive searches of non-sensitive attribute values for a large set of data due to complexity reasons. Therefore, we propose informative search as the strategy S, and use it for the rest of the paper. This is where the power of the Bayesian network can be exploited. We follow the causal links up or down, from the sensitive attribute T — the target attribute (as noted previously, we assume for simplicity that T is only one node) in order to intelligently and efficiently block data. Of course, this search depends upon the choice of B_n. To be more specific: in Bayesian networks, the parent (child), which is the immediate ancestor (descendant), of a target attribute is the set of attributes denoted as $\mathbf{P}$ ($\mathbf{C}$). It can be shown, based upon Markov independence and a conditional entropy inequality, that any ancestor attribute is less informative to the target attribute than is the joined parent attributes in terms of entropy measure. This property is also applicable to a descendant node provided it has no connections with ancestor nodes of T, that do not go through T. According to Bayesian network theories, the parent attributes are analyzed jointly. The search starts with the attributes that are most relevant to the target attribute and stops if the change reaches the specified tolerance

level U. If the values of parent or child attributes are not available for modification because they themselves are sensitive, the search will proceed with parent and child nodes derived from the immediate family. In general, our method of preventing Low from inferring a sensitive attribute value, t_i, involves changing not one, but all, relationships from $\mathbf{P}$ ($\mathbf{C}$) to T, i.e., $Pr(t_i|\mathbf{P_l}), l = 1, .., |\mathbf{P}|$ ($Pr(t_i|\mathbf{C_k}), k = 1, .., |\mathbf{C}|$), where $|\mathbf{P}|$ ($|\mathbf{C}|$) is the size of the support of $\mathbf{P}$ ($\mathbf{C}$). Could this change lead to conflicting results as one relationship decreases and the other one increases in strength? It can be shown that appropriate modifications will not increase those relevant probabilistic relationships.

5 *V, 0* - Data Selection Criterion

In the above section, we described a strategy to determine which nodes to investigate for data blocking. The following section describes a method for determining which data values to change, given that specific nodes have already been chosen by the search strategy S. What criterion is used to select non-sensitive attribute values for modification (blocking)? We choose attribute values which maximally change the probability of target values, $T = t_i$, with respect to the criterion V (Let X_k denote the attribute values, excluding $T = t_i$, of the kth data item of the support of $T = t_i$.) :

$$V \stackrel{\text{def}}{=} \sum_i \sum_k \mid Pr(T = t_i^k | X_k, B_n^H) \; - \; Pr(T = t_i^k | X_k^m, B_n^{H*}) \mid,$$

where X_k^m is X_k with respect to D_H^m. $Pr(T = t_i^k | X_k^m, B_n^{H*})$ is computed as the average over the different B_n^H corresponding to all possible instantiations (see prior discussion about B_n^{L*}). As discussed, τ lets us measure how the functionality of the database for the Low user, after blocking modifications, has changed with respect to Low. Our goal is to

$$\text{Maximize } V, \text{ while keeping } \tau \leq U.$$

Let N denote the total number of attribute values to be modified. Assume that $N = 2$, and that $U = 2\%$. Therefore, we wish to find the placement of two "?" in Table 2, which both maximizes V while still keeping $\tau \leq .02$. The choice that maximizes V is that of blocking the "hepatitis" value for data item 3, and the "depression" value for data item 4. With this blocking, we have that $\log Pr(D_L|B_n^L)$=-54.53, and $\log Pr(D_L^m|B_n^{L*})$=-55.54. Therefore, $\tau = .019 < .02$. Thus, this is the modification that lessens sensitive inference without unacceptably harming performance. Since $\tau = .019 < .02$, can we increase N to 3? The answer is no. When we determine the D_L^m that maximizes V with three blocked values, we have the result that $\tau > .02$. Therefore, we have found the optimal method of modifying D_L, and thus have lessened the inference within the desired performance bounds. Thus, Table 3 is what High should send "down" to Low. The set O is simply the database as modified by blocking.

6 *E* - Post Evaluation

A Bayesian network structure is not written in the heavens. It is a practical construct based upon statistical properties. Therefore, we must have some way to see if we accomplished what we wished to achieve with respect to lessening inferences about sensitive

Table 3: D_L^m — *modified medical records*
(U: uid; H: hepatitis; D: depression; A: AIDS; T: low thyroid; F: transfusion)

U	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
H	n	y	?	y	n	n	y	y	y	n	n	y	n	y	n	y	n	y	n	y
D	n	y	y	?	y	n	n	n	y	n	y	n	n	y	y	n	y	y	y	n
A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T	n	y	n	y	n	n	n	y	n	n	y	n	n	y	y	n	y	n	y	n
F	n	n	y	n	n	n	n	y	n	n	y	n	n	n	n	n	n	y	n	y

information. Therefore, we need to measure the effectiveness of the modification with the modified database and examine the change in probabilities of sensitive data on an individual basis. If modified values of an attribute can be inferred from its probabilistically relevant attributes, values of those relevant attributes are subject to modification as well [13]. We call this the database *ramification* problem. If we have not affected the probabilities sufficiently, we must re-address our choice of the tolerance level. We are continuing to examine methods from Knowledge Discovery and Datamining (KDD) to accomplish this, e.g., C4.5 [15] is useful, provided the sensitive information is not distributed over multiple nodes.

7 Conclusion

The database inference problem has been intensively studied by researchers from academia, government (e.g., health and medical, IRS, Census Bureau) and industry (e.g., internet retailers) in recent years. In this paper, we characterized the database inference prevention system as having six basic elements. This characterization facilitates formal analysis of the database inference and the sensitive data protection problem. We discussed the database inference problem based on the proposed framework, using techniques founded upon the same techniques as that of Bayesian network theory and KDD. Finally, we demonstrated that our approach provides a practical method of lessening database inference.

Acknowledgments

We are grateful to Ruth Heilizer and the anonymous referees for their helpful comments.

References

[1] Pearl, J. (1989) *Probabilistic Reasoning in Intelligent Systems*, Morgan Kauffman.

[2] Spiegelhalter, D. & Lauritzen, S. (1990) "Sequential updating of conditional probabilities on directed graphical structures," Networks, 20, pp, 579-605.

[3] Heckerman, D. (1996) "Bayesian Networks for Knowledge Discovery," Advances in Knowledge Discovery and Data Mining, AAAI Press/MIT Press, pp. 273-305.

[4] Pearl, J. (2000) *Causality*, Cambridge.

[5] Denning, D. & Neumann, P. (1985) "Requirements and Model for IDES-A Real-Time Intrusion-Detection Expert System," # 83F83-01-00 CS SRI International.

[6] Duncan, G. (1995) "Restricted Data versus Restricted Access," In Seminar on New Directions in Statistical Methodology, OMB, pp, 43-56.

[7] Hinke, T., Delugach, H. & Wolf, R. (1997) "Protecting Databases from Inference Attack," Computers & Security, Vol. 16, No. 8, pp, 687-708.

[8] Lin, T.Y., Hinke, T.H., Marks, D.G., & Thuraisingham, B. (1996) "Security and Data Mining," Proc. IFIP WG11.3, *Database Security Vol. 9: Status and Prospects,*.

[9] Zayatz, L. & Rowland, S. (1999) "Disclosure Limitation for American Factfinder," Census Bureau report (manuscript).

[10] Moskowitz, I. S. & Chang, L. (1999) "A Formal View of the Database Inference Problem," Proc. CIMCA'99, pp. 254-259, Feb. 1999, Vienna, Austria.

[11] Moskowitz, I. S. & Chang, L. (2000) "A Computational Intelligence Approach to the Database Inference Problem," *Advances in Intelligent Systems:Theory and Applications* (ed M. Mohammadian) IOS Press.

[12] Chang, L. & Moskowitz, I. S. (1998) "Bayesian Methods Applied to the Database Inference Problem," *Database Security Vol. 12* (ed. Jajodia), pp. 237-251, Kluwer.

[13] Chang, L & Moskowitz, I. S. (2000) "An Integrated Framework for Database Inference and Privacy Protection," Proc. of IFIP WG11.3, The Netherlands, *Database Security Vol. 14*, Kluwer.

[14] Spirtes, P., Glymour, C. and Scheines, R. (1993) *Causation, Prediction, and Search.* Springer-Verlag, NY.

[15] Quinlan, R. (1992) *C4.5*, Morgan Kaufmann.